arXiv:1205.6654v1 [astro-ph.IM] 30 May 2012

Reduced Ambiguity Calibration for LOFAR

Sarod Yatawatta

Abstract Interferometric calibration always yields non unique solutions. It is therefore essential to remove these ambiguities before the solutions could be used in any further modeling of the sky, the instrument or propagation effects such as the ionosphere. We present a method for LOFAR calibration which does not yield a unitary ambiguity, especially under ionospheric distortions. We also present exact ambiguities we get in our solutions, in closed form. Casting this as an optimization problem, we also present conditions for this approach to work. The proposed method enables us to use the solutions obtained via calibration for further modeling of instrumental and propagation effects. We provide extensive simulation results on the performance of our method. Moreover, we also give cases where due to degeneracy, this method fails to perform as expected and in such cases, we suggest exploiting diversity in time, space and frequency.

1 Introduction

Self calibration is essential for radio interferometers such as LOFAR[1] to obtain high quality results under the presence of corruptions. In this paper, we focus most of our attention on propagation effects caused by the ionosphere. The ionosphere is an active medium which affects electromagnetic radiation passing through it. Naturally, radio interferometric observations on the earth are distorted by its effects due to this reason. Most significant distortions were first observed in very long baseline interferometry (VLBI) due to the fact that extremely long baselines will see the sky through unrelated parts of the ionosphere. We refer the reader to Cotton (1995), (and references therein) for an overview of earliest attempts of VLBI calibration under the influence of the

ASTRON, Dwingeloo, NL
E-mail: yatawatta@astron.nl

[1] The Low Frequency Array: http://www.lofar.org

ionosphere. Such attempts have become simpler since the introduction of the matrix measurement equation (Hamaker et al 1996), which encapsulates the full polarized form of radio interferometric data in a concise way.

Ionospheric distortions can also become significant at very low frequencies, even with baselines of moderate length. Therefore, instruments such as LOFAR, MWA, PAPER, GMRT, and SKA which can observe at frequencies close to ionospheric cutoff will certainly have to cope with such disturbances. Significant work has already being done in calibration and correction for ionospheric distortions of interferometric observations at very low frequency and the reader is referred to e.g., (Intema et al 2009; van der Tol and van der Veen 2007). This paper is not about tackling such distortions, but rather about extraction of ionospheric (as well as instrumental) information with minimal ambiguities. As shown by Hamaker (2000), self calibration always introduces certain ambiguities of the solutions. At this point, we shall clarify the difference between a degeneracy and an ambiguity. A degeneracy arises when we do not have enough observations to extract complete information from the data. For example, consider locating a point in three dimensional space, just by distance measurements from a set of known reference points. If we have only two distance measurements, the location of the point will have a degeneracy. That is, it could lie anywhere on the intersection of the two spheres (with radii equal to the measured distances) from the reference points. However, by having three or more properly placed reference points, we could eliminate this degeneracy. In contrast, an ambiguity arises due to our description of a problem, or due to the choice of a coordinate system to describe our parameters. Take the case of finding the square root of a given number: We know that we could always have more than one (complex) solution to this problem. Note that this cannot be eliminated by taking more measurements.

In the case of radio interferometry, we have more than enough baselines compared to the number of stations (or unknowns) to eliminate any degeneracy. However, the ambiguities remain, regardless of the number of baselines or data points we have. This is not a major concern in most cases with an unpolarized sky as the ambiguity will not affect the end result. On the other hand, to construct models for instrumental and propagation effects we need to use solutions obtained by calibration. In such situations, it is essential that we use ambiguity free solutions. Traditional radio telescopes such as the Giant Meterwave Radio Telescope (GMRT) and the Very Long Baseline Array (VLBA) eliminate the ambiguities by periodically using a noise injection source but this if far from practical for a telescope like LOFAR.

Calibration of a phased array interferometer such as LOFAR has its own challenges as well as advantages in comparison with traditional dish based (movable) interferometers. Due to not having any movable components, approaching electromagnetic radiation from a celestial source will have a time varying relationship with the LOFAR dipoles because of the rotation of the sky. On the other hand, due to the same reason (no movable parts), the LOFAR beamshape is simpler to simulate using numerical electromagnetic software. In contrast, a steerable dish based telescope can track a source across the sky but the beamshape is more complex to simulate due to shadowing, spillover etc. Numerical simulation of LOFAR element beamshape has given us extensive knowledge of its behavior. Therefore, in this paper, we shall use

the a priori knowledge of the LOFAR beamshape to our fullest advantage to restrict ambiguities in calibration.

This paper is organized as follows: In section 2, we give a brief overview of LOFAR calibration and the arising ambiguities. Next, in section 3, we derive expressions for the ambiguities in closed form. In section 4, we discuss the necessary conditions to obtain reduced ambiguity solutions and cases where these conditions will not hold. Finally, we present results based on simulations in section 5 before concluding.

Notation: Matrices are denoted in bold uppercase letters (e.g. $\mathbf{A}$) and vectors in bold lowercase letters (e.g. $\mathbf{v}$). The sets of Complex, Real and Integer numbers are denoted as $\mathbb{C}, \mathbb{R}$ and $\mathbb{Z}$ respectively. An estimate (or a possible solution) for a given quantity (e.g. $\boldsymbol{\theta}$) is given by $\widehat{\boldsymbol{\theta}}$. The column vector with all ones is denoted by $\mathbf{1}$ and all zeros by $\mathbf{0}$. The submatrix created by selecting a subset of rows (or columns) from matrix $\mathbf{A}$ is represented as $\mathbf{A}_{[a:b,c:d]}$, where we have selected rows a to b and columns c to d. The pseudoinverse of $\mathbf{A}$ is denoted by $\mathbf{A}^{\dagger}$ while the norm by $\|\mathbf{A}\|$.

2 Calibration

Consider the correlations observed on baseline p-q: We can express the observed visibility matrix (ignoring additive noise) for baseline p-q in the standard (Hamaker et al 1996) Jones formalism as

$$\mathbf{V}_{pq} = \mathbf{G}_p \mathbf{E}_p \mathbf{F}_p \mathbf{C} \mathbf{F}_q^H \mathbf{E}_q^H \mathbf{G}_q^H \tag{1}$$

where we have N stations and (p, q) are drawn from the set

$$\mathcal{S} = \{(1, 2), (1, 3), \ldots, (1, N), \ldots, (N - 1, N)\}. \tag{2}$$

The number of baselines (cardinality of $\mathcal{S}$) is $N_b \le N(N-1)/2$. The matrices $\mathbf{G}, \mathbf{E}$ and $\mathbf{F}$ in (1) correspond to the instrumental gain, element beamshape and the ionospheric Faraday rotation, (qualified by subscript p or q to denote the station) respectively.

We first make the following assumptions, mainly to simplify our analysis:

- We observe a dominant, unpolarized, point source, at the phase center (therefore no Fourier phase component) and the station beams are pointing at that source (therefore station beam gain is unity and $\mathbf{E}$ represents the dipole element beam).
- We assume electronic leakage is small (therefore $\mathbf{G}$ is diagonal). Moreover, if there is any leakage it is assumed to be part of the dipole $\mathbf{E}$ Jones matrix.
- We assume we have a priori knowledge of the $\mathbf{E}$ Jones matrix. Moreover, because all stations have parallel dipoles (as in LOFAR), the $\mathbf{E}$ Jones matrices are assumed to be identical for all stations.
- Ionospheric delay phase is absorbed into the clock delays in $\mathbf{G}$ Jones.

Note that the assumptions made here are specific for a typical LOFAR observation and therefore the results based on these assumptions are mainly applicable to a telescope that is very similar to LOFAR.

The unknowns for station p are given in $\mathbf{G}_p$ and $\mathbf{F}_p$:

$$\mathbf{G}_p = \begin{bmatrix} g_{xp} & 0 \\ 0 & g_{yp} \end{bmatrix}, \ \mathbf{F}_p = \begin{bmatrix} \cos(\alpha_p) & \sin(\alpha_p) \\ -\sin(\alpha_p) & \cos(\alpha_p) \end{bmatrix}, \tag{3}$$

where $g_{xp}, g_{yp} \in \mathbb{C}$ and $\alpha_p \in \mathbb{R}$ are the unknown station gain and Faraday rotation parameters, respectively. The element beam shape determines the $\mathbf{E}$ ($= \mathbf{E}_p = \mathbf{E}_q$) Jones matrices, about which we have fairly accurate knowledge from numerical simulations. The source coherency (with intensity I) is given as

$$\mathbf{C} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}. \tag{4}$$

Consider the problem of calibration: We can represent (1) as

$$\mathbf{V}_{pq} = \mathbf{J}_p \mathbf{C} \mathbf{J}_q^H \tag{5}$$

where $\mathbf{J}_p$ and $\mathbf{J}_q$ are the Jones matrices that we need to estimate. However, we can get solutions for (5) such as $\widehat{\mathbf{J}}_p = \mathbf{J}_p \mathbf{U}$, where $\mathbf{U}$ is an unknown unitary matrix. This is perfectly fine if we only need to correct the data using these solutions because the unitary ambiguity cancels out in the final product. However, if we need to study the parameters in (3), it is hard to disentangle the unitary ambiguity from the solutions.

Another way of looking at this problem is by counting the number of degrees of freedom that is actually present in the problem. If we consider the unknowns g_{xp}, g_{yp} (4 real numbers), α_p (1 real number), we get 5 degrees of freedom per station. However, if we attempt calibration as in (5), we have 8 degrees of freedom, which is clearly an overkill.

Therefore, instead of solving for a system as in (5), we use the original form (1) with unknown parameters as given in (3). Still, this does not guarantee us an ambiguity free solution as we see later. Hence, we study the possible ambiguities of (1) in section 3 and give the necessary conditions to get a solution in section 4.

3 Ambiguity Analysis

Consider the unknowns g_{xp}, g_{yp} in (3). We can rewrite this as $g_{xp} = |g_{xp}| \exp(j\phi_{xp})$, $g_{yp} = |g_{yp}| \exp(j\phi_{yp})$ where ϕ_{xp}, ϕ_{yp} are the unknown (true) phases. In order to study the ambiguities we could get for the unknowns in (3), we could rewrite the possible solutions as

$$\widehat{\mathbf{G}}_p = \begin{bmatrix} (-1)^{k_{xp}}|g_{xp}| \times & 0 \\ \exp\left(j(\widehat{\phi}_{xp} + k_{xp}\pi + k_{\alpha p}\pi)\right) & \\ & (-1)^{k_{yp}}|g_{yp}| \times \\ 0 & \exp\left(j(\widehat{\phi}_{yp} + k_{yp}\pi + k_{\alpha p}\pi)\right) \end{bmatrix}, \tag{6}$$

and

$$\widehat{\mathbf{F}}_p = \begin{bmatrix} \cos(\widehat{\alpha}_p + k_{\alpha p}\pi) & \sin(\widehat{\alpha}_p + k_{\alpha p}\pi) \\ -\sin(\widehat{\alpha}_p + k_{\alpha p}\pi) & \cos(\widehat{\alpha}_p + k_{\alpha p}\pi) \end{bmatrix}, \tag{7}$$

where $k_{xp}, k_{yp}, k_{\alpha p} \in \mathbb{Z}$. Note that if we substitute (6) and (7) instead of (3) into (1), we should still get the same value for $\mathbf{V}_{pq}$. The necessary conditions for us to rewrite (3) as (6) and (7) will be discussed in section 4.

Consider the product

$$\mathbf{E}_{pq} \stackrel{\triangle}{=} \mathbf{E}_p \mathbf{F}_p \mathbf{C} \mathbf{F}_q^H \mathbf{E}_q^H = \begin{bmatrix} E_{11,pq} & E_{12,pq} \\ E_{21,pq} & E_{22,pq} \end{bmatrix} \tag{8}$$

such that $\mathbf{V}_{pq} = \mathbf{G}_p \mathbf{E}_{pq} \mathbf{G}_q^H$. Assume $\mathbf{E}_{pq}$ to have non zero off diagonal elements i.e., $E_{12,pq} \neq 0$, $E_{21,pq} \neq 0$. The necessary conditions for this is related to the same conditions for (6) and (7) to be valid solutions to (1) and will be given in section 4.

Then, in order to get the same $\mathbf{V}_{pq}$ in (1), by substitution of $\mathbf{G}_p, \mathbf{G}_q$ given in (3) as well as $\widehat{\mathbf{G}}_p, \widehat{\mathbf{G}}_q$ given in (6) to (1), we get the following conditions:

$$k_{xp} + k_{yp} = 0, 2, \ldots \tag{9}$$
$$k_{xp} + k_{yq} = 0, 2, \ldots$$
$$k_{yp} + k_{xq} = 0, 2, \ldots$$
$$k_{yp} + k_{yq} = 0, 2, \ldots$$

and

$$\widehat{\phi}_{xp} - \widehat{\phi}_{xq} + k_{xp}\pi - k_{xq}\pi + k_{\alpha p}\pi - k_{\alpha q}\pi = \phi_{xp} - \phi_{xq} \tag{10}$$
$$\widehat{\phi}_{xp} - \widehat{\phi}_{yq} + k_{xp}\pi - k_{yq}\pi + k_{\alpha p}\pi - k_{\alpha q}\pi = \phi_{xp} - \phi_{yq}$$
$$\widehat{\phi}_{yp} - \widehat{\phi}_{xq} + k_{yp}\pi - k_{xq}\pi + k_{\alpha p}\pi - k_{\alpha q}\pi = \phi_{yp} - \phi_{xq}$$
$$\widehat{\phi}_{yp} - \widehat{\phi}_{yq} + k_{yp}\pi - k_{yq}\pi + k_{\alpha p}\pi - k_{\alpha q}\pi = \phi_{yp} - \phi_{yq}.$$

We can write similar condition for all other baselines.

Accumulating all possible constraints such as (9) for all baselines, we have

$$\underbrace{\mathbf{A_k}}_{4N_b \times 2N} \underbrace{\mathbf{k}}_{2N \times 1} = \underbrace{\mathbf{n}}_{4N_b \times 1} \tag{11}$$

where $\mathbf{k} \stackrel{\triangle}{=} [k_{x1}, k_{y1}, k_{x2}, k_{y2} \ldots, k_{yN}]^T$. The vector $\mathbf{n}$ has only even integers as entities. This is a linear Diophantine equation (because $\mathbf{k} \in \mathbb{Z}^{2N}$). The matrix $\mathbf{A}_k$ has ones at two locations on each row. We can show this has rank of $2N$ (full column rank) using similar techniques as in (Chen and Petropulu 2001; Yatawatta et al 2004). Therefore, we get a unique solution to $\mathbf{k}$ (ideally equal to $\mathbf{0}$).

Next, we can also accumulate equations in (10) for all possible baselines to arrive at

$$\underbrace{\mathbf{A_p}}_{4N_b \times 2N} \left(\underbrace{\widehat{\boldsymbol{\phi}}}_{2N \times 1} + \underbrace{\mathbf{k}}_{2N \times 1} \pi + \underbrace{\mathbf{k}_\alpha}_{2N \times 1} \pi \right) = \underbrace{\mathbf{A_p}}_{4N_b \times 2N} \underbrace{\boldsymbol{\phi}}_{2N \times 1} \tag{12}$$

where $\widehat{\boldsymbol{\phi}} \stackrel{\triangle}{=} [\widehat{\phi}_{x1}, \widehat{\phi}_{y1}, \widehat{\phi}_{x2}, \widehat{\phi}_{y2}, \ldots, \widehat{\phi}_{Ny}]^T$, $\boldsymbol{\phi} \stackrel{\triangle}{=} [\phi_{x1}, \phi_{y1}, \phi_{x2}, \phi_{y2}, \ldots, \phi_{Ny}]^T$, and $\mathbf{k}_\alpha \stackrel{\triangle}{=} [k_{\alpha 1}, k_{\alpha 1}, k_{\alpha 2}, k_{\alpha 2}, \ldots, k_{\alpha N}]^T$. The matrix $\mathbf{A}_p$ has 1 and -1 at two locations on each row. Similar as in (Chen and Petropulu 2001; Yatawatta et al 2004), we can show that $\mathbf{A}_p$ has rank $2N - 1$ (null space of rank 1). Therefore, in order to find a solution

to (12), we have to fix one element in $\widehat{\boldsymbol{\phi}}$, say by making $\widehat{\boldsymbol{\phi}}_{[1]} = \widehat{\phi}_{x1} = \phi_0$. Then, we can rewrite (12) as

$$
\mathbf{A}_{p[:,2:2N]} \left(\widehat{\boldsymbol{\phi}}_{[2:2N]} + \mathbf{k}_{[2:2N]}\pi + \mathbf{k}_{\alpha[2:2N]}\pi \right) \tag{13}
$$
$$
= \mathbf{A}_{p[:,2:2N]}\boldsymbol{\phi}_{[2:2N]} - \mathbf{A}_{p[:,1]} \left(-\boldsymbol{\phi}_{[1]} + \phi_0 + \mathbf{k}_{[1]}\pi + \mathbf{k}_{\alpha[1]}\pi \right)
$$

which gives us the solution

$$
\widehat{\boldsymbol{\phi}}_{[2:2N]} = \boldsymbol{\phi}_{[2:2N]} \tag{14}
$$
$$
- \mathbf{A}^{\dagger}_{p[:,2:2N]}\mathbf{A}_{p[:,1]} \left(-\boldsymbol{\phi}_{[1]} + \phi_0 + \mathbf{k}_{[1]}\pi + \mathbf{k}_{\alpha[1]}\pi \right)
$$
$$
- \left(\mathbf{k}_{[2:2N]}\pi + \mathbf{k}_{\alpha[2:2N]}\pi \right).
$$

To simplify (14), we need to find $\mathbf{A}^{\dagger}_{p[:,2:2N]}\mathbf{A}_{p[:,1]}$. We know that the sum of each row of $\mathbf{A}_p$ is zero. Therefore,

$$
\mathbf{A}_p \mathbf{1} = \mathbf{0} \tag{15}
$$
$$
\mathbf{A}_{p[:,2:2N]}\mathbf{1} + \mathbf{A}_{p[:,1]}\mathbf{1} = \mathbf{0}
$$
$$
\mathbf{A}^{\dagger}_{p[:,2:2N]}\mathbf{A}_{p[:,1]} = -\mathbf{1}.
$$

Therefore, we can rewrite (14) as

$$
\widehat{\boldsymbol{\phi}}_{[2:2N]} = \boldsymbol{\phi}_{[2:2N]} + \mathbf{1} \left(-\boldsymbol{\phi}_{[1]} + \phi_0 + \mathbf{k}_{[1]}\pi + \mathbf{k}_{\alpha[1]}\pi \right) \tag{16}
$$
$$
- \left(\mathbf{k}_{[2:2N]}\pi + \mathbf{k}_{\alpha[2:2N]}\pi \right).
$$

Closer scrutiny of (16) reveals that the estimated phase, $\widehat{\boldsymbol{\phi}}_{[2:2N]}$ is equal to the true phase, $\boldsymbol{\phi}_{[2:2N]}$ with a constant unknown phase (common to all stations) and an ambiguity of $k\pi$, where $k \in \mathbb{Z}$. Therefore, the ambiguity of the solutions for the X and Y phases for a given station is the same (within π). The difference between the X phase and Y phase ambiguities is called the phase zero difference in traditional calibration jargon. Thus, we get no phase zero difference (except an ambiguity in π) for all the stations.

Next, we shall investigate the conditions for $\mathbf{F}_p, \mathbf{F}_q$ and $\widehat{\mathbf{F}}_p, \widehat{\mathbf{F}}_q$ to satisfy (1) to get the same $\mathbf{V}_{pq}$. We arrive at equations similar to

$$
\widehat{\alpha}_p - \widehat{\alpha}_q + k_{\alpha p}\pi - k_{\alpha q}\pi = \alpha_p - \alpha_q \tag{17}
$$

for each baseline. Combining all possible equations (17) for all baselines, we get

$$
\underbrace{\mathbf{A}_f}_{N_b \times N} \left(\underbrace{\widehat{\boldsymbol{\alpha}}}_{N \times 1} + \underbrace{\mathbf{k}_{\tilde{\alpha}}}_{N \times 1} \pi \right) = \underbrace{\mathbf{A}_f}_{N_b \times N} \underbrace{\boldsymbol{\alpha}}_{N \times 1} \tag{18}
$$

where $\widehat{\boldsymbol{\alpha}} \triangleq [\widehat{\alpha}_1, \widehat{\alpha}_2, \dots, \widehat{\alpha}_N]^T$, $\boldsymbol{\alpha} \triangleq [\alpha_1, \alpha_2, \dots, \alpha_N]^T$ and $\mathbf{k}_{\tilde{\alpha}} \triangleq [k_{\alpha 1}, k_{\alpha 2}, \dots, k_{\alpha N}]^T$. The matrix $\mathbf{A}_f$ has 1 and -1 at two locations on each row. We can once again show

that the rank of this matrix is $N - 1$. Hence, in order to get a solution, we fix one element of $\widehat{\alpha}$, say $\widehat{\alpha}_{[1]} = \widehat{\alpha}_1 = \alpha_0$. Then, we can rewrite (18) as

$$\mathbf{A}_{f[:,2:N]} \left(\widehat{\alpha}_{[2:N]} + \mathbf{k}_{\tilde{\alpha}[2:N]}\pi \right) = \mathbf{A}_{f[:,2:N]} \alpha_{[2:N]} \tag{19}$$
$$- \mathbf{A}_{f[:,1]} \left(-\alpha_{[1]} + \alpha_0 + \mathbf{k}_{\tilde{\alpha}[1]}\pi \right)$$

to yield a solution as

$$\widehat{\alpha}_{[2:N]} = \alpha_{[2:N]} - \mathbf{A}_{f[:,2:N]}^{\dagger} \mathbf{A}_{f[:,1]} \left(-\alpha_{[1]} + \alpha_0 + \mathbf{k}_{\tilde{\alpha}[1]}\pi \right) - \mathbf{k}_{\tilde{\alpha}[2:N]}\pi. \tag{20}$$

Once again, because the sum of each row in $\mathbf{A}_f$ is zero, using similar analysis as (15), we can show that $\mathbf{A}_{f[:,2:N]}^{\dagger} \mathbf{A}_{f[:,1]} = -\mathbf{1}$. Therefore, we can simplify the solution as

$$\widehat{\alpha}_{[2:N]} = \alpha_{[2:N]} + \mathbf{1} \left(-\alpha_{[1]} + \alpha_0 + \mathbf{k}_{\tilde{\alpha}[1]}\pi \right) - \mathbf{k}_{\tilde{\alpha}[2:N]}\pi. \tag{21}$$

Therefore, our estimate $\widehat{\alpha}_{[2:N]}$ is equal to the true values, $\alpha_{[2:N]}$ plus an unknown constant which is common to all solutions with an ambiguity of $k\pi$, $k \in \mathbb{Z}$. Thus, we can estimate the differential Faraday rotation within a common unknown rotation and an integer ambiguity of π.

4 Necessary Conditions

In this section, we give necessary conditions for our approach to work, especially with respect to the properties of the element beam pattern or the $\mathbf{E}$ Jones matrix.

4.1 Non diagonal $\mathbf{E}_{pq}$ in (8)

First, we study the requirements for (8) to be non diagonal. We express the $\mathbf{E}$ and $\mathbf{F}$ Jones matrices in (8) as

$$\mathbf{E}_p = \mathbf{E}_q = \begin{bmatrix} E_{x\theta} & E_{x\phi} \\ E_{y\theta} & E_{y\phi} \end{bmatrix}, \tag{22}$$

$$\mathbf{F}_p \mathbf{C} \mathbf{F}_q^H = I \begin{bmatrix} \cos(\alpha_p - \alpha_q) & \sin(\alpha_p - \alpha_q) \\ -\sin(\alpha_p - \alpha_q) & \cos(\alpha_p - \alpha_q) \end{bmatrix},$$

where $E_{x\theta}, E_{x\phi}, E_{y\theta}, E_{y\phi}$ are the field components in the directions θ, ϕ in cylindrical polar coordinates. The conditions required for the off diagonal terms to be zero then becomes:

$$\cos(\alpha_p - \alpha_q)(E_{x\theta}E_{y\theta}^{\star} + E_{x\phi}E_{y\phi}^{\star}) \tag{23}$$
$$+ \sin(\alpha_p - \alpha_q)(E_{x\theta}E_{y\phi}^{\star} - E_{x\phi}E_{y\theta}^{\star}) = 0$$
$$\cos(\alpha_p - \alpha_q)(E_{x\theta}^{\star}E_{y\theta} + E_{x\phi}^{\star}E_{y\phi})$$
$$+ \sin(\alpha_p - \alpha_q)(-E_{x\theta}^{\star}E_{y\phi} + E_{x\phi}^{\star}E_{y\theta}) = 0.$$

Note also that the off diagonal terms of the product $\mathbf{E}\mathbf{E}^H$ is equal to $E_{x\theta}E_{y\theta}^{\star} + E_{x\phi}E_{y\phi}^{\star}$ (and its conjugate). If this term is zero, we get a solution to (23) as $\alpha_p = \alpha_q + 2k\pi$.

On the other hand, if $E_{x\theta}E_{y\phi}^{\star} - E_{x\phi}E_{y\theta}^{\star} = 0$, then we get another solution to (23) as $\alpha_p = \alpha_q + (2k + 1)\pi/2$. So, if either of these conditions are satisfied by the element beam, our method fails to give a unique solution.

A more general way of looking at (23) is as follows: We can represent (23) in matrix form as

$$\widetilde{\mathbf{E}}\mathbf{c} = \mathbf{0} \tag{24}$$

where

$$\widetilde{\mathbf{E}} \triangleq \begin{bmatrix} (E_{x\theta}E_{y\theta}^{\star} + E_{x\phi}E_{y\phi}^{\star}) & (E_{x\theta}E_{y\phi}^{\star} - E_{x\phi}E_{y\theta}^{\star}) \\ (E_{x\theta}E_{y\theta}^{\star} + E_{x\phi}E_{y\phi}^{\star})^{\star} & -(E_{x\theta}E_{y\phi}^{\star} - E_{x\phi}E_{y\theta}^{\star})^{\star} \end{bmatrix}, \tag{25}$$

$$\mathbf{c} \triangleq \begin{bmatrix} \cos(\alpha_p - \alpha_q) \\ \sin(\alpha_p - \alpha_q) \end{bmatrix}.$$

We can find a solution for $\mathbf{c}$ in (24) if at least one of the eigenvalues of $\widetilde{\mathbf{E}}$ is zero (and the solution for $\mathbf{c}$ is the scaled eigenvector). Therefore, when $\widetilde{\mathbf{E}}$ has at least one zero eigenvalue, it is possible (under certain values of Faraday rotation) that our method fails. On the other hand, when none of the eigenvalues are zero, we are guaranteed that regardless of the Faraday rotation, our method works.

4.2 Convex Optimization

Normally, the solutions for the calibration are found using a non linear optimization technique. Therefore, its natural to consider the solution of (1) as an optimization problem. The residual error of (1) can be given as

$$\mathbf{r}_{pq} = vec(\mathbf{V}_{pq}) - \mathbf{f}_{pq}(\boldsymbol{\theta}) \tag{26}$$

where $\mathbf{f}_{pq}(\boldsymbol{\theta}) = vec\left(\mathbf{G}_p\mathbf{E}_p\mathbf{F}_p\mathbf{C}\mathbf{F}_q^H\mathbf{E}_q^H\mathbf{G}_q^H\right)$. The unknown parameter vector is given by $\boldsymbol{\theta}$. For each station we have 5 (real) unknowns, i.e., $|g_{xp}|, |g_{yp}|, \phi_{xp}, \phi_{yp}$ and α_p. However, since we can only find differential Faraday rotation (not the absolute rotation), we keep α for one station fixed. So, the parameter vector can be given as $\boldsymbol{\theta} \triangleq [|g_{x1}|, |g_{x2}|, \ldots, |g_{y1}|, |g_{y2}|, \ldots, \phi_{x1}, \phi_{x2}, \ldots, \phi_{y1}, \phi_{y2}, \ldots, \alpha_2, \alpha_3, \ldots]^T$. Hence, for N stations, the size of $\boldsymbol{\theta}$ is $5N - 1$. Finally, we can write the calibration problem as

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\theta} \sum_{p,q} \mathbf{r}_{pq}^H \mathbf{r}_{pq}. \tag{27}$$

The Hessian of this optimization problem will have rank $5N - 2$ at the solution. This is because, we still have an ambiguity in ϕ_{xp} or ϕ_{yp}. Therefore, the smallest eigenvalue of the Hessian will be zero (or close to zero under noise). For a well formulated optimization problem, we need to show that (27) is convex (Boyd and Vendenberghe 2004) with respect to $\boldsymbol{\theta}$. This implies that the Hessian should be positive (semi) definite. In other words, all eigenvalues (barring the smallest eigenvalue which is zero) should be positive. Therefore, by studying the second smallest eigenvalue, we can study the performance of our calibration, in particular with variation of the beam in

different directions of the sky. The more positive this is, the more well formulated our problem becomes.

Under noise, the Cramer-Rao bound is inversely proportional to the second smallest eigenvalue that we discussed. Therefore, we can study the same behavior by using the Hessian instead of the Cramer-Rao bound in this case.

4.3 Fréchet Derivative and Condition Number

We have assumed a priori knowledge of $\mathbf{E}$ in our analysis. However, there will certainly be small variations in the real beamshape from our assumed model. We need to show that small changes in $\mathbf{E}$ in (1) only leads to small changes in $\mathbf{V}$. This ensures that the proposed method works, even when the real $\mathbf{E}$ will be slightly different from our model and moreover, when each station will have a slightly different beamshape.

We use the Fréchet derivative of (1) to examine the sensitivity of $\mathbf{V}$ to small changes in $\mathbf{E}$. The Fréchet derivative is a generalization of the *derivative* of scalar functions to higher dimensional functional spaces. More details about the Fréchet derivative and sensitivity analysis of matrix functions can be found in Higham (2008). The dependence of $\mathbf{E}$ on $\mathbf{V}_{pq}$ in (1) can be written as

$$\mathbf{V}(\mathbf{E}) = \mathbf{V}_{pq} = \mathbf{G}_p \mathbf{E} \mathbf{F}_p \mathbf{C} \mathbf{F}_q^H \mathbf{E}^H \mathbf{G}_q^H \tag{28}$$

Let $\mathbf{\Delta E}$ denote a small change to $\mathbf{E}$. Then, as in Higham (2008), we write the Fréchet derivative as

$$L_{\mathbf{V}}(\mathbf{E}, \mathbf{\Delta E}) = \mathbf{V}(\mathbf{E} + \mathbf{\Delta E}) - \mathbf{V}(\mathbf{E}) - o(\|\mathbf{\Delta E}\|) \tag{29}$$

where

$$\lim_{\|\mathbf{\Delta E}\| \to 0} \frac{o(\|\mathbf{\Delta E}\|)}{\|\mathbf{\Delta E}\|} = 0. \tag{30}$$

Therefore, we get the Fréchet derivative as

$$L_{\mathbf{V}}(\mathbf{E}, \mathbf{\Delta E}) = \mathbf{G}_p \mathbf{E} \mathbf{F}_p \mathbf{C} \mathbf{F}_q^H \mathbf{\Delta E}^H \mathbf{G}_q^H + \mathbf{G}_p \mathbf{\Delta E} \mathbf{F}_p \mathbf{C} \mathbf{F}_q^H \mathbf{E}^H \mathbf{G}_q^H. \tag{31}$$

The relative condition number of $\mathbf{V}(\mathbf{E})$ is given as

$$cond(\mathbf{V}(\mathbf{E})) = \max_{\|\mathbf{\Delta E}\| \neq 0} \frac{\|L_{\mathbf{V}}(\mathbf{E}, \mathbf{\Delta E})\|}{\|\mathbf{\Delta E}\|} \frac{\|\mathbf{E}\|}{\|\mathbf{V}(\mathbf{E})\|}. \tag{32}$$

Using norm inequalities $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$ and $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$ we get an upper bound for the relative condition number as

$$cond(\mathbf{V}(\mathbf{E})) \leq \|\mathbf{G}_p\| (\|\mathbf{F}_p \mathbf{C} \mathbf{F}_q^H \mathbf{E}^H\| + \|\mathbf{E} \mathbf{F}_p \mathbf{C} \mathbf{F}_q^H\|) \|\mathbf{G}_q\| \frac{\|\mathbf{E}\|}{\|\mathbf{V}(\mathbf{E})\|}. \tag{33}$$

We can use (33) to study the sensitivity of the data to variations in $\mathbf{E}$ and hence, the sensitivity of our calibration to errors in our knowledge of $\mathbf{E}$. Generally, it is better to have a condition number as small as possible, the smallest being equal to 1.

4.4 Diversity

In this section, we have given several ways of looking at the necessary conditions for our approach to work. In practice, there will always be cases where some of these conditions are not satisfied. However, even if these conditions are not satisfied, we can exploit diversity to make our method work. Diversity can be described as the opposite of degeneracy. Indeed, degeneracy arises when we do not have enough independent observations. In contrast, when we do have observations taken independently from one another we create diversity. There are several different forms of diversity that we can exploit. Implicitly, we used the knowledge of $\mathbf{E}$ and moreover, the fact that this is not an identity matrix, in our proposed method. In other words, we used polarization diversity arising due to X and Y dipoles having different beam patterns. Furthermore, even if the aforementioned conditions are not satisfied for one baseline, there are other baselines where the ionosphere is different and a solution is possible. Secondly, ionosphere will almost always change with time and even for a baseline which has degeneracy, with time, things will change to yield a solution. Thirdly, the element beam changes with direction in the sky and as the source moves across the beam, it will overcome degenerate cases. Moreover, ionospheric properties have well known frequency behavior. In situations where we fail to obtain a solution, we can extrapolate solutions obtained at different times,baselines,frequencies etc., to get a complete solution.

5 Simulation Results

We present some results based on numerical simulations in this section. We select the observation frequency to be at 120 MHz. The $\mathbf{E}$ Jones matrix is derived from numerical electromagnetic simulation of a LOFAR high band antenna (HBA) dipole. The elements of the $\mathbf{E}$ Jones matrix at this frequency is shown in Fig. 1 and Fig. 2. Both these figures show the beam over the full hemisphere, orthographically projected onto the plane. The dipoles are rotated by $\pi/4$ radians from the meridian. The center of each figure correspond to the zenith. The perimeter of the figure correspond to the horizon. Note that in all figures in this section we use the same projection and coordinate system (except in Fig. 6 (b)). In Fig. 3, we have also shown the power beam for both X and Y dipoles.

Now, let us reconsider the necessary conditions discussed in section 4. In Fig. 4, we have shown the conditions when (23) could be satisfied. We have shown the products $|(E_{x\theta}E_{y\theta}^{\star} + E_{x\phi}E_{y\phi}^{\star})|$ and $|(E_{x\theta}E_{y\phi}^{\star} - E_{x\phi}E_{y\theta}^{\star})|$ in Fig. 4 (a) and (b), respectively. The first term goes close to zero along the planes of the dipoles and at the horizon. However, the second term only goes to zero at the horizon.

In Fig. 4 (c), we have shown magnitude of the smallest eigenvalue of $\widetilde{\mathbf{E}}$ (normalized by $\|\widetilde{\mathbf{E}}\|$) given in (25). This figure tells us something more than the previous two figures: which is that apart from the directions along the planes of the dipoles (which also includes the zenith), no possible Faraday rotation exists such that (23) is satisfied. In other words, regardless of the Faraday rotation, we should be able to estimate
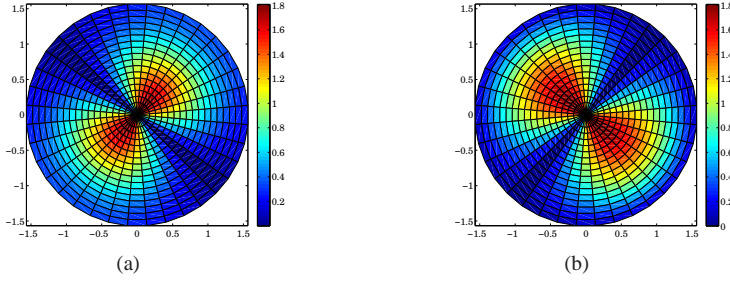
Fig. 1 The X beam amplitude at 120 MHz. (a) amplitude $|E_{x\theta}|$ (b) amplitude $|E_{x\phi}|$
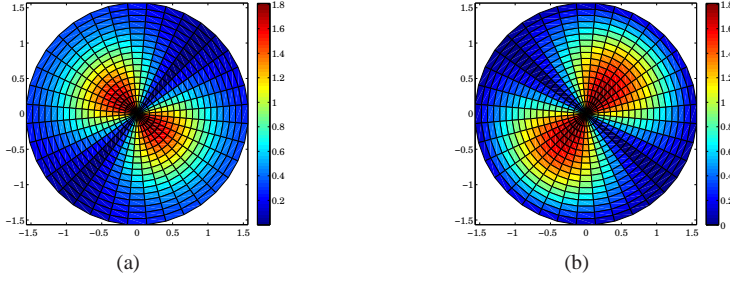

Fig. 2 The Y beam amplitude at 120 MHz. (a) amplitude $|E_{y\theta}|$ (b) amplitude $|E_{y\phi}|$
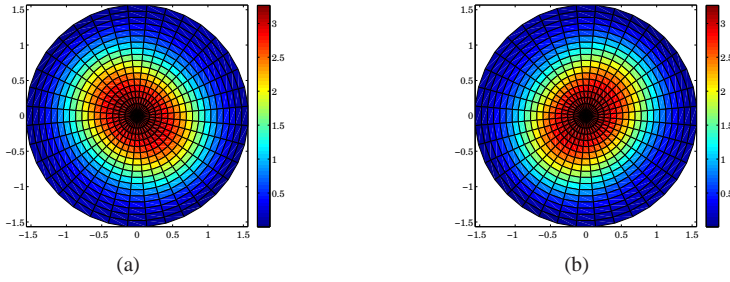
Fig. 3 The power beam for X and Y dipoles at 120 MHz. (a) X power beam $|E_{x\theta}|^2 + |E_{x\phi}|^2$ (b) Y power beam $|E_{y\theta}|^2 + |E_{y\phi}|^2$

it in all directions except the planes of the dipole and the horizon, because (8) is not diagonal.

So far, we have studied the characteristics of the beam that enables our calibration to work. Now, we shall consider the full optimization problem and evaluate its performance. We have simulated an array with $N = 54$ LOFAR stations. We vary the direction of the celestial source in the sky (keeping the flux at 1 Jy), and for each direction, we do a simulation. The direction is varied at discrete intervals with azimuth range $[0, 2\pi)$ radians and elevation range $[0, \pi/2]$ radians. The gains $|g_{xp}|, |g_{yp}|$ are simulated from a uniform distribution in $[0.5, 1.5]$. The phases ϕ_{xp}, ϕ_{yp} are simu-
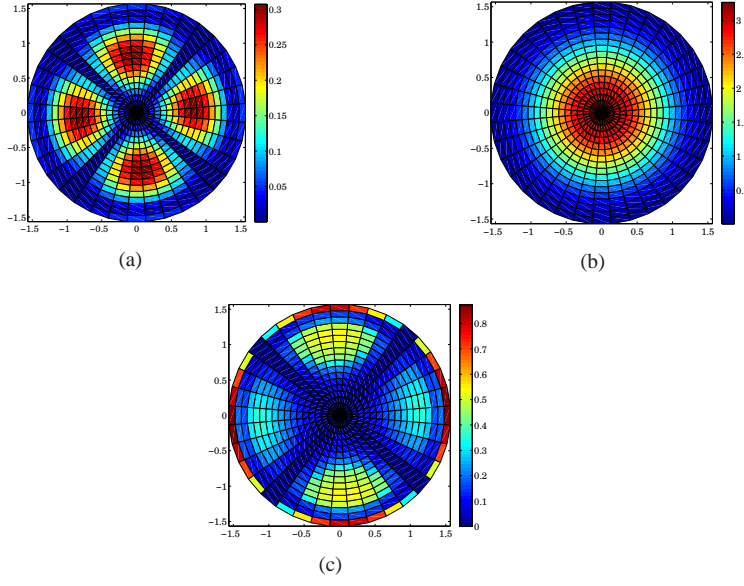
Fig. 4 Conditions for (8) to be not diagonal. (a) amplitude $|(E_{x\theta}E_{y\theta}^{\star} + E_{x\phi}E_{y\phi}^{\star})|$ (b) amplitude $|(E_{x\theta}E_{y\phi}^{\star} - E_{x\phi}E_{y\theta}^{\star})|$ (c) amplitude of smallest eigenvalue of $\widetilde{\mathbf{E}}$

lated from a uniform distribution in $[0, 2\pi)$ radians. The Faraday rotations α_p are also simulated from a uniform distribution in $[0, 2\pi)$ radians. We have also added white Gaussian noise to the simulated visibilities, keeping the signal to noise ratio at 10. We use a stand alone optimization routine (Lourakis 2004) to solve the calibration problem (27). The initial values for $|g_{xp}|, |g_{yp}|$ are kept at 1 and the rest of parameters are kept at 0.
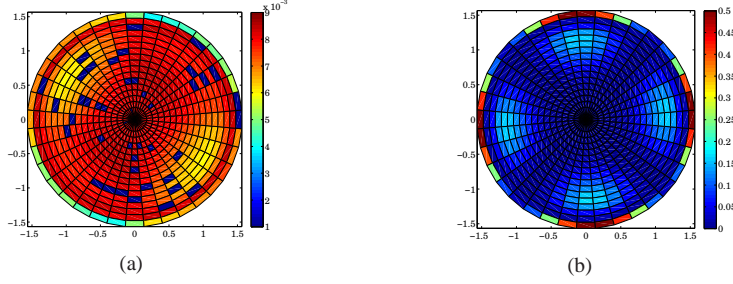

Fig. 5 (a) The second smallest eigenvalue of the Hessian of (27). (b) Residual normalized error of (27).

In Fig. 5 (a), we have shown the second smallest eigenvalue of the Hessian obtained after calibration. This is positive for most directions in the sky and where it goes close to zero, we need to exploit diversity to obtain a complete solution. How-

ever, the overall behavior shows that the problem is mostly convex. In Fig. 5 (b), we have shown the normalized residual error of (27), which is small and mostly dominated by noise, except for some high values at the horizon.

In Fig 6 (a), we have shown the relative condition number obtained using (33). The average value of the condition number over all baselines is shown. It is below 3 for this particular case, indicating that the problem is well conditioned and quite robust to error in modeling of $\mathbf{E}$. Next, in Fig. 6 (b), we have shown the histogram of the errors in estimating ϕ_{xp} and ϕ_{yp}. The true error is small and hardly visible in this figure. The dominant feature is the ambiguity in π. We see similar behavior in the solutions for α_p.

(a)	(b)
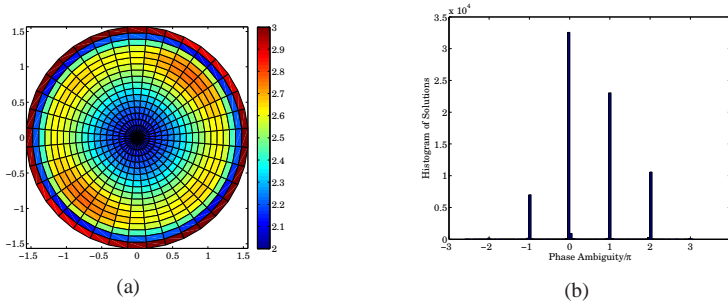
Fig. 6 (a) Relative condition number (33). (b) Histogram of the error of the solutions for ϕ_{xp} and ϕ_{yp}, divided by π.

6 Conclusions

We have shown that it is indeed possible to calibrate LOFAR with reduced ambiguities. The analysis was done for an unpolarized point source, and it is straight forward to extend this to an unpolarized extended source. The crucial factor in our analysis was that the a priori knowledge of the element beam shape. Extensive electromagnetic simulations has in fact given us that knowledge. We have also shown under which conditions our calibration fails to give an answer. However, in such situations, we can exploit diversity to get a complete solution. The extension of this method to polarized sources as well as results based on application of the proposed method to real LOFAR data will be presented in future work.

7 Acknowledgments

References

Boyd S, Vendenberghe L (2004) Convex optimization. Cambridge U.K.:Cambridge University Press

Chen B, Petropulu AP (2001) Frequency domain blind MIMO system identification based on second- and higher order statistics. IEEE Trans on Signal Processing 49(1):1677–1688

Cotton WD (1995) Polarimetry. ASP Conference Series, Very Long Baseline Interferometry and the VLBA, JA Zensus, PJ Diamond and PJ Napier (eds) 82:289–308

Hamaker JP (2000) Understanding Radio Polarimetry IV: The full-coherency analogue of scalar selfcalibration. Astronomy and Astrophysics Supp 143(3):515–534

Hamaker JP, Bregman JD, Sault RJ (1996) Understanding radio polarimetry, paper I. Astronomy and Astrophysics Supp 117(137):96–109

Higham NJ (2008) Functions of matrices: Theory and computation. Philadelphia USA: SIAM

Intema HT, van der Tol S, Cotton WD, Cohen AS, van Bemmel IM, Rottgering HJA (2009) onospheric calibration of low frequency radio interferometric observations using the peeling scheme. I. Method description and first results. A&A 501(3):1185–1205

Lourakis MIA (2004) levmar: Levenberg-Marquardt nonlinear least squares algorithms in C/C++. `http://www.ics.forth.gr/~lourakis/levmar/`

van der Tol S, van der Veen AJ (2007) Ionospheric Calibration for the LOFAR Radio Telescope. IEEE International Symposium on Signals, Circuits and Systems, 2007 ISSCS 2007 pp 1–4

Yatawatta S, Petropulu AP, Dattani R (2004) Blind channel estimation using fractional sampling. IEEE Trans on Vehicular Technology 53(2):363–371